



ENTER

TECH CENTER 1600/2900

MAR 1 8 2002

RECEIVED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/840,008

DATE: 03/05/2002

TIME: 13:59:07

Input Set : A:\Salk2270.app

Output Set: N:\CRF3\03052002\I840008.raw

```

3 <110> APPLICANT: EVANS, RONALD M.
5 <120> TITLE OF INVENTION: XENOBIOTIC COMPOUND MODULATED EXPRESSION SYSTEMS AND
6   USES THEREFOR
8 <130> FILE REFERENCE: SALK2270-4
10 <140> CURRENT APPLICATION NUMBER: 09/840,008
11 <141> CURRENT FILING DATE: 2001-04-20
13 <150> PRIOR APPLICATION NUMBER: 09/458,366
14 <151> PRIOR FILING DATE: 1999-12-09
16 <150> PRIOR APPLICATION NUMBER: 09/005,286
17 <151> PRIOR FILING DATE: 1998-01-09
19 <160> NUMBER OF SEQ ID NOS: 43
21 <170> SOFTWARE: PatentIn Ver. 2.1
23 <210> SEQ ID NO: 1
24 <211> LENGTH: 2068
25 <212> TYPE: DNA
26 <213> ORGANISM: Homo sapiens
28 <220> FEATURE:
29 <221> NAME/KEY: CDS
30 <222> LOCATION: (583)..(1884)
32 <220> FEATURE:
33 <221> NAME/KEY: modified_base
34 <222> LOCATION: (1263)
35 <223> OTHER INFORMATION: a, c, t, or g
37 <400> SEQUENCE: 1
38 ggcacgagga gatctagggt caaattaatg ttgccctag ttgtaaagga cagagaccct 60
40 cagactgatg aaatgcgctc agaattactt agacaaagcg gatatttgcc actctcttcc 120
42 ccttttcctg tgtttttgta gtgaagagac ctgaaagaaa aaagtaggga gaacataatg 180
44 agaacaaata cggtaatctc ttcatttgct agttcaagtg ctggacttgg gacttaggag 240
46 gggcaatgga gccgcttagt gcctacatct gacttggact gaaatatagg tgagagacaa 300
48 gattgtctca tatccgggga aatcataacc tatgactagg acgggaagag gaagcactgc 360
50 ctttacttca gtgggaatct cggcctcagc ctgcaagcca agtggttcaca gtgagaaaag 420
52 caagagaata agctaatact cctgtcctga acaaggcagc ggctccttgg taaagctact 480
54 ccttgatcga tcctttgcac cggattgttc aaagtggacc ccaggggaga agtcggagca 540
56 aagaacttac caccaagcag tccaagaggc ccagaagcaa ac ctg gag gtg aga 594
57                                     Leu Glu Val Arg
58                                     1
60 ccc aaa gaa agc tgg aac cat gct gac ttt gta cac tgt gag gac aca 642
61 Pro Lys Glu Ser Trp Asn His Ala Asp Phe Val His Cys Glu Asp Thr
62   5          10          15          20
64 gag tct gtt cct gga aag ccc agt gtc aac gca gat gag gaa gtc gga 690
65 Glu Ser Val Pro Gly Lys Pro Ser Val Asn Ala Asp Glu Glu Val Gly
66          25          30          35
68 ggt ccc caa atc tgc cgt gta tgt ggg gac aag gcc act ggc tat cac 738

```

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/840,008

TIME: 13:59:07

Input Set : A:\Salk2270.app

Output Set: N:\CRF3\03052002\I840008.raw

```

69 Gly Pro Gln Ile Cys Arg Val Cys Gly Asp Lys Ala Thr Gly Tyr His
70          40          45          50
72 ttc aat gtc atg aca tgt gaa gga tgc aag ggc ttt ttc agg agg gcc 786
73 Phe Asn Val Met Thr Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg Ala
74          55          60          65
76 atg aaa cgc aac gcc cgg ctg agg tgc ccc ttc cgg aag ggc gcc tgc 834
77 Met Lys Arg Asn Ala Arg Leu Arg Cys Pro Phe Arg Lys Gly Ala Cys
78          70          75          80
80 gag atc acc cgg aag acc cgg cga cag tgc cag gcc tgc cgc ctg cgc 882
81 Glu Ile Thr Arg Lys Thr Arg Arg Gln Cys Gln Ala Cys Arg Leu Arg
82 85          90          95          100
84 aag tgc ctg gag agc ggc atg aag aag gag atg atc atg tcc gac gag 930
85 Lys Cys Leu Glu Ser Gly Met Lys Lys Glu Met Ile Met Ser Asp Glu
86          105          110          115
88 gcc gtg gag gag agg cgg gcc ttg atc aag cgg aag aaa agt gaa cgg 978
89 Ala Val Glu Glu Arg Arg Ala Leu Ile Lys Arg Lys Lys Ser Glu Arg
90          120          125          130
92 aca ggg act cag cca ctg gga gtg cag ggg ctg aca gag gag cag cgg 1026
93 Thr Gly Thr Gln Pro Leu Gly Val Gln Gly Leu Thr Glu Glu Gln Arg
94          135          140          145
96 atg atg atc agg gag ctg atg gac gct cag atg aaa acc ttt gac act 1074
97 Met Met Ile Arg Glu Leu Met Asp Ala Gln Met Lys Thr Phe Asp Thr
98          150          155          160
100 acc ttc tcc cat ttc aag aat ttc cgg ctg cca ggg gtg ctt agc agt 1122
101 Thr Phe Ser His Phe Lys Asn Phe Arg Leu Pro Gly Val Leu Ser Ser
102 165          170          175          180
104 ggc tgc gag ttg cca gag cct ctg cag gcc cca tcg agg gaa gaa gct 1170
105 Gly Cys Glu Leu Pro Glu Pro Leu Gln Ala Pro Ser Arg Glu Glu Ala
106          185          190          195
108 gcc aag tgg agc cag gtc cgg aaa gat ctg tgc tct ttg aag gtc tct 1218
109 Ala Lys Trp Ser Gln Val Arg Lys Asp Leu Cys Ser Leu Lys Val Ser
110          200          205          210
W--> 112 ctg caa gct gcg ggg gga gga tgg cag tgt ctg gaa cta caa acn ccc 1266
W--> 113 Leu Gln Ala Ala Gly Gly Gly Trp Gln Cys Leu Glu Leu Gln Xaa Pro
114          215          220          225
116 agc cga cag tgg cgg aaa gag atc ttc tcc ctg ctg ccc cac atg gct 1314
117 Ser Arg Gln Trp Arg Lys Glu Ile Phe Ser Leu Leu Pro His Met Ala
118          230          235          240
120 gac atg tca acc tac atg ttc aaa ggc atc atc agc ttt gcc aaa gtc 1362
121 Asp Met Ser Thr Tyr Met Phe Lys Gly Ile Ile Ser Phe Ala Lys Val
122 245          250          255          260
124 atc tcc tac ttc agg gac ttg ccc atc gag gac cag atc tcc ctg ctg 1410
125 Ile Ser Tyr Phe Arg Asp Leu Pro Ile Glu Asp Gln Ile Ser Leu Leu
126          265          270          275
128 aag ggg gcc gct ttc gag ctg tgt caa ctg aga ttc aac aca gtg ttc 1458
129 Lys Gly Ala Ala Phe Glu Leu Cys Gln Leu Arg Phe Asn Thr Val Phe
130          280          285          290
132 aac gcg gag act gga acc tgg gag tgt ggc cgg ctg tcc tac tgc ttg 1506
133 Asn Ala Glu Thr Gly Thr Trp Glu Cys Gly Arg Leu Ser Tyr Cys Leu

```

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/840,008

TIME: 13:59:07

Input Set : A:\Salk2270.app

Output Set: N:\CRF3\03052002\I840008.raw

```

134          295          300          305
136 gaa gac act gca ggt ggc ttc cag caa ctt cta ctg gag ccc atg ctg 1554
137 Glu Asp Thr Ala Gly Gly Phe Gln Gln Leu Leu Leu Glu Pro Met Leu
138          310          315          320
140 aaa ttc cac tac atg ctg aag aag ctg cag ctg cat gag gag gag tat 1602
141 Lys Phe His Tyr Met Leu Lys Lys Leu Gln Leu His Glu Glu Glu Tyr
142 325          330          335          340
144 gtg ctg atg cag gcc atc tcc ctc ttc tcc cca gac cgc cca ggt gtg 1650
145 Val Leu Met Gln Ala Ile Ser Leu Phe Ser Pro Asp Arg Pro Gly Val
146          345          350          355
148 ctg cag cac cgc gtg gtg gac cag ctg cag gag caa ttc gcc att act 1698
149 Leu Gln His Arg Val Val Asp Gln Leu Gln Glu Gln Phe Ala Ile Thr
150          360          365          370
152 ctg aag tcc tac att gaa tgc aat cgg ccc cag cct gct cat agg ttc 1746
153 Leu Lys Ser Tyr Ile Glu Cys Asn Arg Pro Gln Pro Ala His Arg Phe
154          375          380          385
156 ttg ttc ctg aag atc atg gct atg ctc acc gag ctc cgc agc atc aat 1794
157 Leu Phe Leu Lys Ile Met Ala Met Leu Thr Glu Leu Arg Ser Ile Asn
158          390          395          400
160 gct cag cac acc cag cgg ctg ctg cgc atc cag gac ata cac ccc ttt 1842
161 Ala Gln His Thr Gln Arg Leu Leu Arg Ile Gln Asp Ile His Pro Phe
162 405          410          415          420
164 gct acg ccc ctc atg cag gag ttg ttc ggc atc aca ggt agc 1884
165 Ala Thr Pro Leu Met Gln Glu Leu Phe Gly Ile Thr Gly Ser
166          425          430
168 tgagcggctg ccttgggtga caccttcgag aggcagccag acccagagcc ctctgagccg 1944
170 gcactcccgg gccaaagacag atggacactg ccaagagccg acaatgccct gctggcctgt 2004
172 ctccctaggg aattcctgct atgacagctg gctagcattc ctcaggaagg acatgggggtg 2064
174 cccc 2068
177 <210> SEQ ID NO: 2
178 <211> LENGTH: 434
179 <212> TYPE: PRT
180 <213> ORGANISM: Homo sapiens
182 <220> FEATURE:
183 <221> NAME/KEY: MOD_RES
184 <222> LOCATION: (227)
185 <223> OTHER INFORMATION: Threonine
187 <400> SEQUENCE: 2
188 Leu Glu Val Arg Pro Lys Glu Ser Trp Asn His Ala Asp Phe Val His
189 1 5 10 15
191 Cys Glu Asp Thr Glu Ser Val Pro Gly Lys Pro Ser Val Asn Ala Asp
192 20 25 30
194 Glu Glu Val Gly Gly Pro Gln Ile Cys Arg Val Cys Gly Asp Lys Ala
195 35 40 45
197 Thr Gly Tyr His Phe Asn Val Met Thr Cys Glu Gly Cys Lys Gly Phe
198 50 55 60
200 Phe Arg Arg Ala Met Lys Arg Asn Ala Arg Leu Arg Cys Pro Phe Arg
201 65 70 75 80
203 Lys Gly Ala Cys Glu Ile Thr Arg Lys Thr Arg Arg Gln Cys Gln Ala

```

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/840,008

TIME: 13:59:07

Input Set : A:\Salk2270.app

Output Set: N:\CRF3\03052002\I840008.raw

```

204          85          90          95
206 Cys Arg Leu Arg Lys Cys Leu Glu Ser Gly Met Lys Lys Glu Met Ile
207          100          105          110
209 Met Ser Asp Glu Ala Val Glu Glu Arg Arg Ala Leu Ile Lys Arg Lys
210          115          120          125
212 Lys Ser Glu Arg Thr Gly Thr Gln Pro Leu Gly Val Gln Gly Leu Thr
213          130          135          140
215 Glu Glu Gln Arg Met Met Ile Arg Glu Leu Met Asp Ala Gln Met Lys
216 145          150          155          160
218 Thr Phe Asp Thr Thr Phe Ser His Phe Lys Asn Phe Arg Leu Pro Gly
219          165          170          175
221 Val Leu Ser Ser Gly Cys Glu Leu Pro Glu Pro Leu Gln Ala Pro Ser
222          180          185          190
224 Arg Glu Glu Ala Ala Lys Trp Ser Gln Val Arg Lys Asp Leu Cys Ser
225          195          200          205
227 Leu Lys Val Ser Leu Gln Ala Ala Gly Gly Gly Trp Gln Cys Leu Glu
228          210          215          220
W--> 230 Leu Gln Xaa Pro Ser Arg Gln Trp Arg Lys Glu Ile Phe Ser Leu Leu
231 225          230          235          240
233 Pro His Met Ala Asp Met Ser Thr Tyr Met Phe Lys Gly Ile Ile Ser
234          245          250          255
236 Phe Ala Lys Val Ile Ser Tyr Phe Arg Asp Leu Pro Ile Glu Asp Gln
237          260          265          270
239 Ile Ser Leu Leu Lys Gly Ala Ala Phe Glu Leu Cys Gln Leu Arg Phe
240          275          280          285
242 Asn Thr Val Phe Asn Ala Glu Thr Gly Thr Trp Glu Cys Gly Arg Leu
243          290          295          300
245 Ser Tyr Cys Leu Glu Asp Thr Ala Gly Gly Phe Gln Gln Leu Leu Leu
246 305          310          315          320
248 Glu Pro Met Leu Lys Phe His Tyr Met Leu Lys Lys Leu Gln Leu His
249          325          330          335
251 Glu Glu Glu Tyr Val Leu Met Gln Ala Ile Ser Leu Phe Ser Pro Asp
252          340          345          350
254 Arg Pro Gly Val Leu Gln His Arg Val Val Asp Gln Leu Gln Glu Gln
255          355          360          365
257 Phe Ala Ile Thr Leu Lys Ser Tyr Ile Glu Cys Asn Arg Pro Gln Pro
258          370          375          380
260 Ala His Arg Phe Leu Phe Leu Lys Ile Met Ala Met Leu Thr Glu Leu
261 385          390          395          400
263 Arg Ser Ile Asn Ala Gln His Thr Gln Arg Leu Leu Arg Ile Gln Asp
264          405          410          415
266 Ile His Pro Phe Ala Thr Pro Leu Met Gln Glu Leu Phe Gly Ile Thr
267          420          425          430
269 Gly Ser
272 <210> SEQ ID NO: 3
273 <211> LENGTH: 25
274 <212> TYPE: DNA
275 <213> ORGANISM: Artificial Sequence
277 <220> FEATURE:

```

## RAW SEQUENCE LISTING

DATE: 03/05/2002

PATENT APPLICATION: US/09/840,008

TIME: 13:59:07

Input Set : A:\Salk2270.app

Output Set: N:\CRF3\03052002\I840008.raw

```

278 <223> OTHER INFORMATION: Description of Artificial Sequence: Putative SXR
279     response element from the steroid hydroxylase,
280     rCYP3A1
282 <400> SEQUENCE: 3
283 tagacagttc atgaagttca tctac                                25
286 <210> SEQ ID NO: 4
287 <211> LENGTH: 25
288 <212> TYPE: DNA
289 <213> ORGANISM: Artificial Sequence
291 <220> FEATURE:
292 <223> OTHER INFORMATION: Description of Artificial Sequence: Putative SXR
293     response element from the steroid hydroxylase,
294     rCYP3A2
296 <400> SEQUENCE: 4
297 taagcagttc ataaagttca tctac                                25
300 <210> SEQ ID NO: 5
301 <211> LENGTH: 25
302 <212> TYPE: DNA
303 <213> ORGANISM: Artificial Sequence
305 <220> FEATURE:
306 <223> OTHER INFORMATION: Description of Artificial Sequence: Putative SXR
307     response element from the steroid hydroxylase,
308     rUGT1A6
310 <400> SEQUENCE: 5
311 actgtagttc ataaagttca catgg                                25
314 <210> SEQ ID NO: 6
315 <211> LENGTH: 26
316 <212> TYPE: DNA
317 <213> ORGANISM: Artificial Sequence
319 <220> FEATURE:
320 <223> OTHER INFORMATION: Description of Artificial Sequence: Putative SXR
321     response element from the steroid hydroxylase,
322     rbCYP2C1
324 <400> SEQUENCE: 6
325 caatcagttc aacagggttc accaat                                26
328 <210> SEQ ID NO: 7
329 <211> LENGTH: 33
330 <212> TYPE: DNA
331 <213> ORGANISM: Artificial Sequence
333 <220> FEATURE:
334 <223> OTHER INFORMATION: Description of Artificial Sequence: Putative SXR
335     response element from the steroid hydroxylase,
336     rP450R
338 <400> SEQUENCE: 7
339 cacaggtgag ctgaggccag cagcaggtcg aaa                        33
342 <210> SEQ ID NO: 8
343 <211> LENGTH: 27
344 <212> TYPE: DNA
345 <213> ORGANISM: Artificial Sequence

```

→ Use of n and / or Xaa has been detected in the Sequence Listing. Review the Sequence Listing to ensure a corresponding explanation is present in the <220> to <223> fields of each sequence using n or Xaa.

## VERIFICATION SUMMARY

DATE: 03/05/2002

PATENT APPLICATION: US/09/840,008

TIME: 13:59:08

Input Set : A:\Salk2270.app

Output Set: N:\CRF3\03052002\I840008.raw

L:112 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:113 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:230 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2  
L:545 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:22  
L:564 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:23